I, Tadahiko Itoh, a Patent Attorney of Tokyo, Japan having my office at 32nd Floor, Yebisu Garden Place Tower, 20-3 Ebisu 4-Chome, Shibuya-Ku, Tokyo 150-6032, Japan do solemnly and sincerely declare that I am the translator of the attached English language translation and certify that the attached English language translation is a correct, true and faithful translation of Japanese Patent Application No. 10-251193 to the best of my knowledge and belief.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Tadahiko ITOH
Patent Attorney
ITOH International Patent Office
32nd Floor,
Yebisu Garden Place Tower,
20-3 Ebisu 4-Chome, Shibuya-Ku,
Tokyo 150-6032, Japan

JPA No. 10-251193

(Document Name)	Application for Patent
(Reference Number)	NTTH105937
(Date of Submission)	September 4, 1998
(Destination)	Commissioner of Patent Office
(IPC)	G06C
(Title of the Invention)	METHOD AND APPARATUS FOR EXTRACTED DIGITAL WATERMARKING DATA STATISTICAL PROCESSING, AND A PROGRAM STORAGE MEDIUM
(Number of Claims)	16
(Inventor)	
(Residence or Address)	C/O NIPPON TELEGRAPH AND TELEPHONE CORPORATION 19-2 Nishi-Shinjuku 3-chome Shinjuku-ku, Tokyo, Japan
(Name)	Hiroshi Ogawa
(Inventor)	
(Residence or Address)	C/O NIPPON TELEGRAPH AND TELEPHONE CORPORATION 19-2 Nishi-Shinjuku 3-chome Shinjuku-ku, Tokyo, Japan
(Name)	Takao Nakamura
(Inventor)	
(Residence or Address)	C/O NIPPON TELEGRAPH AND TELEPHONE CORPORATION 19-2 Nishi-Shinjuku 3-chome Shinjuku-ku, Tokyo, Japan
(Name)	Atsuki Tomioka
(Inventor)	
(Residence or Address)	C/O NIPPON TELEGRAPH AND TELEPHONE CORPORATION 19-2 Nishi-Shinjuku 3-chome Shinjuku-ku, Tokyo, Japan
(Name)	Youichi Takashima
(Applicant for Patent)	
(Identification Number)	000004226
(Name)	NIPPON TELEGRAPH AND TELEPHONE CORPORATION
(Attorney)	
(Identification Number)	100066153
(Patent Attorney)	
(Name)	Takashi Kusano

JPA No. 10-251193

(Designated Attorney)
(Identification Number) 100100642
(Patent Attorney)
(Name) Minoru Inagaki
(Indication of Official Fees)
(Prepayment Ledger Number) 002897
(Amount Paid) ¥21,000
(Lists of Submitted Documents)
(Document Name) Specification 1
(Document Name) Drawing 1
(Document Name) Abstract 1
(Number of General Power of Attorney) 9806848
(Proof Requested or Not) Requested

[Name of the Document] Specification

[Title of the Invention]

METHOD AND APPARATUS FOR EXTRACTED DIGITAL WATERMARKING DATA STATISTICAL PROCESSING, AND A PROGRAM STORAGE MEDIUM

[Claims]

1. An extracted digital watermark data statistical processing method for reconstituting the digital watermark data embedded in data contents, comprising the step of:

reconstituting the digital watermark data embedded in the data contents from a data sequence prior to the reconstitution embedded from the data contents by means of a detection method on the basis of a binary distribution in statistics.

2. The extracted digital watermark data statistical processing method as claimed in claim 1, further comprising the steps of:

presetting a reliability threshold value of the extracted digital watermark data;

calculating an appearance probability of the extracted digital watermark data prior to the reconstitution in the data contents, on the basis of the digital watermark data obtained from the bias of appearance probability of the digital watermark data in the binary distribution of appearance probability of each bit of 1 bit sequence extracted at random from digital data contents; and

reconstituting digital watermark data by using majority decision processing if said appearance probability or 1-appearance probability exceeds threshold value, and determining that there is no watermark or the presence is unknown if said appearance probability or 1-appearance probability does not exceed the threshold value.

3. The extracted digital watermark data statistical processing method as claimed in claim 2, further comprising the step of:
judging from the appearance probability of the extracted digital watermark data prior to the reconstitution in the data contents in order to reconstitutes the extracted digital watermark data.

4. The extracted digital watermark data statistical processing method as claimed any of the claims 1 through 3, further comprising the step of:

obtaining the digital watermark data and its reliability from a bias of the appearance probability of the extracted digital watermark data prior to the reconstitution in the binary distribution obtained from the appearance probability of each bit of 1 bit sequence extracted at random from digital data contents of the appearance probability of the extracted digital watermark data prior to the reconstitution.

5. The extracted digital watermark data statistical processing method as claimed any of the claims 1 through 4, further comprising the steps of:

modulating in advance the data sequence actually embedded as the digital watermark data; and

demodulating said bit sequence by said pseudo-random sequence which is used for embedding the digital watermark prior to the digital data reconstituting processing.

6. The extracted digital watermark data statistical processing method as claimed in the claim 5, further comprising the step of:

fixing both of the each bit of the appearance probability of each bit of 1 bit sequence extracted at random from digital data contents to 1/2.

7. An extracted digital watermark data

statistical processing apparatus for reconstituting the digital watermark data embedded in data contents, comprising:

means for reconstituting the digital watermark data embedded in the data contents from a data sequence prior to the reconstitution embedded from the data contents by means of a detection method on the basis of a binary distribution in statistics.

8. The extracted digital watermark data statistical processing apparatus as claimed in claim 7, wherein said extraction means further comprising

means for obtaining a binary distribution from a digital watermark data length and each bit of the appearance probability of each bit of 1 bit sequence extracted at random from digital data contents;

means for extracting the digital watermark sequence with respect to the each bit of said data contents;

means for obtaining from the appearance probability of the extracted digital watermark data;

means for judging whether the obtained appearance probability is larger the reliability threshold value; and

means for reconstituting the digital watermark data of which the data is judged as larger than said threshold value.

9. The extracted digital watermark data statistical processing apparatus as claimed in claim 8, wherein said extraction means further comprising :

means for obtaining the reliability of said reconstituted data bit from said obtained appearance probability and outputting together with the reconstituted data bit.

10. The extracted digital watermark data statistical processing apparatus as claimed in any of the claims 7 to 9, further comprising:

means for modulating in advance the data sequence actually embedded as the digital watermark data by means of pseudo-random sequence; and

means for demodulating said bit sequence by said pseudo-random sequence which is used for embedding the digital watermark prior to the digital data reconstituting processing.

11. An extracted digital watermark data statistical processing program storage medium for reconstituting the digital watermark data embedded in data contents, comprising:

a computer performance means for the reconstituting processing of the digital watermark data embedded in the data contents from a data sequence prior to the reconstitution embedded from the data contents by means of a detection method on the basis of a binary distribution in statistics.

12. The extracted digital watermark data statistical processing program storage medium as claimed in claim 11, wherein said program further comprising:

processing means for presetting a reliability threshold value of the extracted digital watermark data, and

calculating an appearance probability of the extracted digital watermark data prior to the reconstitution in the data contents, on the basis of the digital watermark data obtained from the bias of appearance probability of the digital watermark data in the binary distribution of appearance probability of each bit of 1 bit sequence extracted at random from digital data contents; and

processing means for reconstituting digital watermark data by using majority decision processing if said appearance probability or 1-appearance probability exceeds threshold value, and determining that there is no watermark or the presence is unknown if said appearance probability or 1-appearance probability does not exceed the threshold value.

13. The extracted digital watermark data statistical processing program storage medium as claimed in claim 12, wherein said program further comprising: processing means for judging from the appearance probability of the extracted digital watermark data prior to the reconstitution in the data contents in order to reconstitutes the extracted digital watermark data.

14. The extracted digital watermark data statistical processing program storage medium as claimed in any of the claims 11 through 13, wherein said program further comprising:

processing means for obtaining the digital watermark data and its reliability from a bias of the appearance probability of the extracted digital watermark data prior to the reconstitution in the binary distribution obtained from the appearance probability of each bit of 1 bit sequence extracted at random from digital data contents of the appearance probability of the extracted digital watermark data prior to the reconstitution.

15. The extracted digital watermark data statistical processing program storage medium as claimed in any of the claims 11 through 14, wherein said program further comprising:

processing means for modulating in advance the data sequence actually embedded as the digital watermark data, and demodulating said bit sequence by said pseudo-random sequence which is used for embedding the digital watermark prior to the digital data reconstituting processing.

16. The extracted digital watermark data statistical processing program storage medium as claimed in claims 15, wherein said program further comprising:

processing means for fixing both of the each bit of the appearance probability of each bit of 1 bit sequence extracted at random from digital data contents to 1/2.

[Detailed Description of the Invention]

[Field of the Invention]

It is easy to replicate or tamper fraudulently with multimedia production, and the easiness hinders an data content provider from sending data. In addition, some users may not use the data originated from the provider validly. Therefore, copyright protection is strongly needed for the multimedia production. The digital watermarking technique is effective in realizing the copyright protection. According to the digital watermarking technique, sub-data is embedded in data contents without being noticed by a user by utilizing redundancy of data such as of an image and a sound. The digital watermarking technique is used for protecting a multimedia copyright by embedding copyright information, a user ID and the like as the sub-data in secret, since it is difficult to separate the sub-data from the data contents.

[Prior Art]

Conventional techniques are proposed in Japanese patent applications No.8-305370, No.8-338769, No.9-9812, No.9-14388, No.9-109924, No.9-197003, No.9-218467 and No.10-33239. The digital watermark method is also called data hiding, finger printing steganography, image/sound deep encryption and the like. As for a digital watermarking system, accuracy for determining the presence or absence of embedded data is important. In addition, reliability of embedded data is important. The digital watermarking system generally has a mechanism for reconstituting correct digital watermark data even when sub-data embedded in the data contents is corrupted to a certain extent, since the digital watermarking system assumes various processing on the watermarked data contents. However, under present circumstances, it is impossible for the system to evaluate validity of reconstituted digital watermark data quantitatively. Therefore, the system does not have enough reliability.

[Object of the Invention]

It is an object of the present invention to evaluate quantitatively probabilities of cases that data contents which do not contain digital watermark data are wrongly judged as containing digital watermark data, and incorrect digital watermark data is read from watermarked digital data contents.

[Means to Solve the Problems]

Digital watermark processing is comprised of a pair of digital watermark embedding/digital watermark extraction. In digital watermark embedding processing, digital watermark embedding area $B \in A$ is selected from the digital watermark target area in data contents by means of secret key information so that data in area B is changed by an inherent rule. In digital watermark extraction processing, data of the digital watermark embedding area B is interpreted and the digital watermark data is reconstituted. The present invention judges the probability of occurrence of the digital watermark data read from the watermark embedding area B by means of the correct secret key data, based on a binary distribution in statistics of digital watermark data read by means of any secret key data regardless of true or false from A which is an overall digital watermark target area, by means of the digital watermark algorism which is to be an applied target of the invention, in the data contents embedded with digital watermark.

Effect

The present invention, in an digital watermark technique, can evaluate the credibility of the watermark data read from the data contents, can judge whether the data contents have the digital watermark or not, and can suppress the probability of reading incorrect digital watermark in an certain value from data contents which the digital watermark is included.

[Embodiments of the Invention]

Embodiment 1

Before explaining embodiments of the present invention, definition of some words will be given. "Digital watermark data sequence" represents a data sequence read from the digital data contents before being reconstituted. "Digital watermark data" represents significant data for system operation, which data needs to be embedded in the digital data contents, or, data obtained by reconstituting the digital watermark sequence. "Reliability α of digital watermark" is an index representing validity of read digital watermark data. That is, it represents a probability that the read digital watermark data matches with the actual embedded digital watermark data. Conversely, a probability of reading digital watermark data from an image without digital watermark data or reading erroneous digital watermark data can be represented as $2(1-\alpha)$.

Similarly, "embedded sequence" represents data to be actually embedded. The embedded sequence includes sequence of embedded data which is modulated, extended or repeated. In addition, "read" may be replaced with "extract" in some cases.

Fig.1 shows a digital watermarking system of the present invention. In the system shown in Fig.1, digital watermark data 101 is embedded in digital data contents 103 by a digital watermark embedding apparatus 102, then, converted into watermarked digital data contents 104.

The watermarked digital data contents 104 are degraded to watermarked digital data contents 105 by compression or image processing while the watermarked digital data contents 104 are distributed by wireless or wire communication or by a packaged medium.

A digital watermark reading apparatus 107 reads a watermark sequence from the degraded watermarked digital data contents 105, and reconstitutes digital watermark data 108.

Fig.2 is a block diagram of the watermark reading apparatus 107. The digital watermark data reconstitution apparatus 108 provided in the watermark reading apparatus 107 obtains the probability q that bit 1 is read when any 1 bit watermark sequence is read from a whole watermark area beforehand by using the watermark reading apparatus 107.

Specifically, assuming a 1 bit watermark sequence reading part 501, the part 501 reads the watermark sequence 1 bit by 1 bit from all elements of the whole watermark area (a broken line L1), and calculates the ratio of the number of bit 1 to the number of all trials.

In the embodiment, the reading probability of bit 1 and the number of bit 1 are obtained. However, it is possible that the reading probability of bit 0 and the number of bit 0 are obtained. Basically, there is no difference between the former and the latter. The difference is only on implementation.

Accordingly, the probabilities of detecting bit 0 and 1 when reading 1 bit at random in the watermark area by using the digital watermarking algorithm is calculated to be 1-q and q respectively.

The n bit watermark sequence reading part 502 reads the digital watermark data sequence from the watermarked digital data contents for the number of total times of embedding digital watermark data.

Here, digital watermark data is defined as b_0, b_1, ..., b_{m-1}, $b_i \in \{0, 1\}$, $i < m$ (m bit length), the repeating number of embedding ith bit of the digital watermark data in the digital data contents is defined as n_i, the read watermark sequence is defined as $b'_{0,0}$, $b'_{0,1}$, ...$b'_{0,n0-1}$, $b'_{1,0}$, $b'_{1,1}$, ... $b'_{1,n1-1}$,... ,$b'_{m-1,0}$, $b'_{m-1,1}$, ... $b'_{m-1,nm-1\ -1}$ $b_{i,j} \in \{0, 1\}$ ($\sum_{r=0}^{m-1} n_r$ bit length).

The data reconstitution apparatus 108 receives a subsequence of the digital watermark data sequence one after another from a subsequence corresponding to 0th digital watermark data to a subsequence corresponding to (m-1)th digital watermark data (a solid line L2).

Next, the method for reconstituting ith bit of the digital watermark data will be described concretely.

When n_i bits of digital watermark data sequence is read at random from the watermark area, the probability $P(x=k)$ of k '1' bits appearing in the n_i bit sequence is represented by the binary distribution density function

$$P(x=k)={}_{n_i}C_k q^k \cdot (1-q)^{ni-k} \qquad (1)$$

and the distribution function of that, $F(x)$, is

$$F(x)= \sum_{k=0}^{x} {}_{n_i}C_k q^k \cdot (1-q)^{ni-k} \quad (0 \leqq x \leqq n_i). \qquad (2)$$

Here, ${}_{n_i}C_k$ is the number of combinations when selecting k out of n_i.

Setting a reliability threshold value α $(1/2<\alpha \leqq 1)$ of the digital watermark data, the number of bit 1 included in a subsequence $b'_{i,0}$, $b'_{i,1}$, $\ldots b'_{i,ni-1}$ corresponding to ith digital watermark data is calculated by

$$k_i = \sum_{r=0}^{ni-1} b'_{i,r}\ .$$

Then, digital watermark data is determined in the following way by using the formula (2):

$$b_i = \begin{cases} 0 & \text{when } 0 \leqq F(k_i) \leqq 1-\alpha \\ 1 & \text{when } \alpha \leqq F(k_i) \leqq 1 \\ \text{unknown or not present} & \text{when } 1-\alpha < F(k_i) < \alpha \end{cases} \qquad (3)$$

Viewing from a different angle, when determining by the number of bit 1 included in the watermark sequence n_i, if the largest integer x_0 that satisfies $0 \leqq F(x = x_0) \leqq 1-\alpha$ and the smallest integer x_1 that satisfies $\alpha \leqq F(x = x_1) \leqq 1$ are assumed to be threshold values, the digital watermark data is judged as shown in Fig.3 such that if the number of 1 in n_i is equal to or smaller than x_0 , the digital watermark data is 0, and that if the number of 1 is equal to or larger than x_1, the digital watermark data is 1.

The horizontal axis of Fig.3 represents the number of bit 1 included in the watermark sequence, and the vertical axis represents frequency of the corresponding number. As for unwatermarked digital data contents, the frequency that bit 1 appears in a bit sequence read at random from the digital data contents becomes binary distribution. Thus, the peak of the frequency is at the half point of the number of bits. On the other hand, as for watermarked contents, in the subsequence n_i in which bit 0 is embedded as digital watermark data, the frequency of bit 1 is 0 if there is no degradation and it is a small number which is equal to or smaller than x_0 even if there is degradation. In the subsequence n_i in which bit 1 is embedded as digital watermark data, the frequency of bit 1 is n1 if there is no degradation and it is a large number which is equal to or larger than x_1 even if there is degradation. In this way, the distribution of the frequency of bit 1 or bit 0 in the watermarked sequence is leaning to one side from the center of the binary distribution. The present invention uses the lean for reconstituting digital watermark data from the read watermark sequence.

Depending on a watermarking system, a following method can be used. That is, reconstituted digital watermark data is obtained by using the bias from the central value of the distribution P(x) of the watermark sequence extracted from digital data contents 105. Next, the probability of appearing the read watermark sequence is calculated by the formula (2). Then, if the reconstituted digital watermark data is 1, $F(k_i)$ can be added to watermark dada as the reliability, and, if the reconstituted digital watermark data is 0, $1-F(k_i)$ can be added. The reliability $F(k_i)$ and $1-F(k_i)$ of the digital watermark data is obtained from the bias of appearance probability of the digital watermark data in the binary distribution of appearance probability of each bit of 1 bit sequence extracted at random from digital data contents.

Fig.4 shows a concept in which the length of the digital watermark data is extended to m bits.

The digital watermark data reconstitution apparatus 108 outputs the reconstituted digital watermark data b_0, b_1, …, b_{m-1} as read digital watermark data 106.

Fig.5 is a flowchart showing the above-mentioned process. The process will be described in the following with reference to Fig.5.

Watermarked digital data contents 105 and key data which is necessary for reading digital watermark data is input, and a digital watermark data sequence is extracted with respect to each bit value in step 1. Then, a threshold value α of the reliability is set in step 2, and a probability q that bit 1 appears when 1 bit of digital watermark data is read at random from the whole watermark area is obtained in step 3. Then, a binary distribution function F(x) which represents probability that x bits of 1 are included in the bit sequence is obtained from the probability q and the repeating number n_i of each bit of digital watermark data in step 4.

Then, 0 is assigned to i which distinguish a subsequence of the digital watermark data sequence in step 5. Next, the number of bit 1 in the subsequence is obtained as $k_i = \sum_{r=0}^{ni-1} b'_{i,r}$ and the appearance probability $F(k_i)$ is obtained, then it is determined whether $F(k_i)$ is equal to or less than $1-\alpha$ in step 6. If $F(k_i) \leqq 1-\alpha$, the digital watermark data w'_i is reconstituted as 0 in step 7. Then, i is incremented by

1 in step 8, and the process goes back to step 6 if i<m in step 9. If $F(k_i) \leqq 1-\alpha$ is not true in step 6, it is checked whether $F(k_i) \geqq \alpha$ is true in step 10. If $F(k_i) \geqq \alpha$, the digital watermark data w_i is reconstituted as 1 in step 11, and the process goes to step 8. If $F(k_i) \geqq \alpha$ is not true in step 10, the process ends by determining as there is no watermark or the presence or absence is unknown in step 12. If i is more than n_i in step 9, a reconstituted watermark sequence $\{w'_i\}$ is output. In the above process, the reading process in step 1 can be carried out between step 4 and step 5. In step 6, it is checked whether $1-F(k_i)$ is more than α.

In the first embodiment, it is assumed that there is no bias in the distribution represented by formula (1), that is, $q \cong 1/2$.

When the embedding number n_i of each bit of digital watermark data is adequate for obtaining a statistical characteristic, it becomes $q \cong 1/2$ generally. However, since the value of q depends on characteristics of an watermarking algorithm and digital data contents, q may take a value deviating largely from 1/2 in some rare cases. A method for solving this problem will be described in a second embodiment.

Second Embodiment

In the following, the second embodiment will be described. Fig.6 is a block diagram of a watermarking system of the second embodiment.

The watermark embedding apparatus 102 embeds digital watermark data 101 in digital data contents 103. At the time, when embedding each bit value n_i times repeatedly, watermark sequence is modulated and embedded in the digital data contents 103. The modulation is carried out by a pseudo-random sequence generator (A) 501 which is provided in the watermark embedding apparatus 102.

For example, when assuming the embedding sequence as $b_{0,0}$, $b_{0,1}$, $\ldots b_{0,n0\ -1}$, $b_{1,0}$, $b_{1,1}$, $\ldots$ $b_{1,n1-1}, \ldots$ $, b_{m-1,0}$, $b_{m-1,1}$, $\ldots b_{m-1,nm-1\ -1}$ $b_{i,j} \in \{0,\ 1\}$, and the pseudo-random sequence as $r_{i,0}$, $r_{i,1}$, $\ldots r_{i,ni-1}$ $b_{i,j} \in \{0,\ 1\}$, the embedding sequence is modulated to

$m_{i,0}$, $m_{i,1}$, $\ldots m_{i,ni-1}$

$m_{i,j} = b_{i,j}$ (+) $r_{i,j}$

by the pseudo-random sequence. A(+)B represents XOR of A and B.

According to the above-mentioned process, the same pseudo-random sequence is necessary for digital watermark data reading.

For example, if 1 bit watermark sequence is read by using an M-sequence as the pseudo-random sequence, it becomes $q \cong 1/2$. Therefore, the present invention can be applicable without depending on the watermarking algorithm and digital data contents.

When digital watermark data reading, demodulation is carried out as $b'_{i,j} = m_{i,j} (+) r_{i,j}$ by using a pseudo-random sequence generator (B) 502 which is provided in the watermark reading apparatus 106.

Here, the pseudo-random sequence generator (A) 501 and the pseudo-random sequence generator (B) 502 needs to be implemented such that both of the generators generate the same pseudo-random sequence.

Watermark data is reconstituted with the method of the first embodiment from the watermark sequence $b'_{0,0}$, $b'_{0,1}$, …$b'_{0,n0-1}$, $b'_{1,0}$, $b'_{1,1}$, … $b'_{1,n1-1}$,… ,$b'_{m-1,0}$, $b'_{m-1,1}$, …$b'_{m-1,nm-1\ -1}$ $b_{i,j} \in \{0, 1\}$ obtained by the demodulation.

Since it is considered that the appearance probability q of bit 1 in the watermark sequence can be approximated by the binary distribution regardless of the presence or absence of modulation, there is no influence on the distribution of the density function (1) due to the modulation shown in this embodiment.

In addition, q=1/2 can be assumed in implementation, that is, no process is necessary for obtaining q. Therefore, the amount of processing that is required for watermark reconstitution thus becomes the same as that for majority decision processing. Thus, the reconstitution process becomes faster.

Third Embodiment

In the following, a third embodiment will be described. In the third embodiment, an example will be described showing concrete values on the basis of the first embodiment and the second embodiment. In this embodiment, it is assumed that digital watermark data is 1 bit, the repeating number n of embedding is 127 and the probability q that bit 1 is read when reading 1 bit watermark sequence at random from the whole watermark area is 1/2. If the threshold value α is 0.99999 (which means 99.999%), x_0 in Fig.21 is 36 and x_1 is 90. That is to say, according to the present invention, under the above-mentioned condition, digital watermark data is judged as bit 0 if the number of '1' appeared in the watermark sequence (n bits) is equal to or less than 36, and it is judged as bit 1 if the number of '1' appeared in the watermark sequence (n bits) is equal to or more than 90, and it is judged that there is no watermark data or the presence or absence is unknown in other cases. If it is judged that there is digital watermark data, the correctness of more than 99.999% can be ensured.

In the following, examples of experiments will be shown. In the following experiments, an image of "lena" which has 128×128 pixels is used as a test image, and the threshold value α of the reliability is assumed to be 0.999999.

First Experiment

In this experiment, 1 bit digital watermark data '1' was embedded 127 times repeatedly using key data '50,000', and the watermark sequence was read with various key data. Fig. 7 shows the number of bit '1' in the read watermark sequence corresponding to the key data. In Fig. 7, the vertical axis shows the number of bit '1' in the read watermark sequence, and the horizontal axis shows the key data value. In this experiment, the appearance frequency of bit '1' in the watermark area A was q=0.492247.

When correct key data (50,000) is used, it is judged that digital watermark data is '1' with 99.9999% correctness since the number of bit '1' is more than the threshold value x_1 for judging the presence of watermark. When incorrect key data is used, it is judged that there is no watermark data or the presence or absence is unknown.

Second Experiment

In the second experiment, a watermark sequence which was modulated with a 7 stage M-sequence (initial state is 64) was embedded, and a similar experiment as the first experiment was carried out with various key data and M-sequences of various initial states. The result is shown in Fig. 8. By carrying out the modulation, the value of q becomes 0.500000 from 0.492247, and the variance becomes 31.718777 from 31.008265. Thus, the values are almost not changed from those of the first experiment. It is only when correct key data and correct pseudo-random sequence are used that digital watermark data can be read. In addition, when the watermark sequence is embedded in half data of the watermark area A, q=0.741547 with the modulation and q=0.499768 without the modulation.

[Advantages of the Invention]

The effects of the present invention corresponding to the second object is as follows.

(1) There are following effects by judging digital watermark data on the basis of the binary distribution in statistics:

- The probabilities of following cases can be evaluated quantitatively. The cases are: digital data contents which do not contain digital watermark data are wrongly judged as containing digital watermark data, and incorrect digital watermark data is read from watermarked digital data contents. In addition, the probability can be suppressed within 2(1-α) by using the reliability threshold α of digital watermark data.

(2) There are following effects by modulating digital watermark data by a pseudo-random sequence before embedding the digital watermark data:

- The bias of the probability q of reading bit '1' when 1 bit watermark sequence is read at random from the whole watermark area.
- It becomes difficult to detect the presence or absence of watermark data and the value from the bias of q without the correct key data and the pseudo-random sequence, the key data being necessary for reading digital watermark data and the pseudo-random sequence being necessary for demodulating read watermark sequence.
- In an implementation, since it can be assumed to be q=1/2, the amount of processing that is required for watermark reconstitution becomes the same as that for majority decision processing. Thus, the speed of the processing becomes higher.

α is an index which represents a lower limit of the correctness rate of read digital watermark data, and is manageable in the digital watermarking system. Therefor, the method of using α is superior to a conventional method of showing the correctness rate of read digital watermark data to a user.

The present invention becomes more effective in combination with an error correction code. That is, when a part of bits in digital watermark data is intensively corrupted, it is judged that only the part of bits is unknown and other bit data is in high correctness rate. Therefore, correct data can be read by correcting only the corrupted bit data.

[Brief Description of the Drawings]

Fig.1 is a diagram showing an overview of a digital watermarking system;

Fig.2 is a diagram showing an overview of digital watermarking extracted apparatus;

Fig.3 is a diagram showing a judgment of digital watermarking data;

Fig.4 is a diagram showing an overview of digital watermarking data reconstitution;

Fig.5 is a flowchart showing steps of a digital watermark extraction process;

Fig.6 is a diagram showing an overview of the second embodiment of this invention;

Fig.7 is a diagram showing result (no modulation) of the watermark sequence readings;

Fig.8 is a diagram showing result (with modulation) of the watermark sequence readings.

[Name of the Document] Abstract

[Abstract]

[Object] Object of this invention is to quantitatively evaluate the probability of reading the incorrect digital watermark from the data contents which include the digital watermark.

[Solution Means] When n_i bits of digital watermark data sequence is read at random from the watermark area, the probability P(x=k) of k '1' bits appearing in the n_i bit sequence is represented by the binary distribution density function

$$P(x=k)={}_{n_i}C_k q^k \cdot (1-q)^{ni-k} \quad (1)$$

and the distribution function of that, F(x), is

$$F(x)= \sum_{k=0}^{x} {}_{n_i}C_k q^k \cdot (1-q)^{ni-k} \quad (0 \leqq x \leqq n_i). \quad (2)$$

Here, ${}_{n_i}C_k$ is the number of combinations when selecting k out of n_i.

Setting a reliability threshold value α $(1/2<\alpha \leqq 1)$ of the digital watermark data,

[Selected Figure] Fig. 3

~~【書類名】~~ ~~図面~~ [Name of Document] DRAWING

FIG. 1 ~~【図1】~~

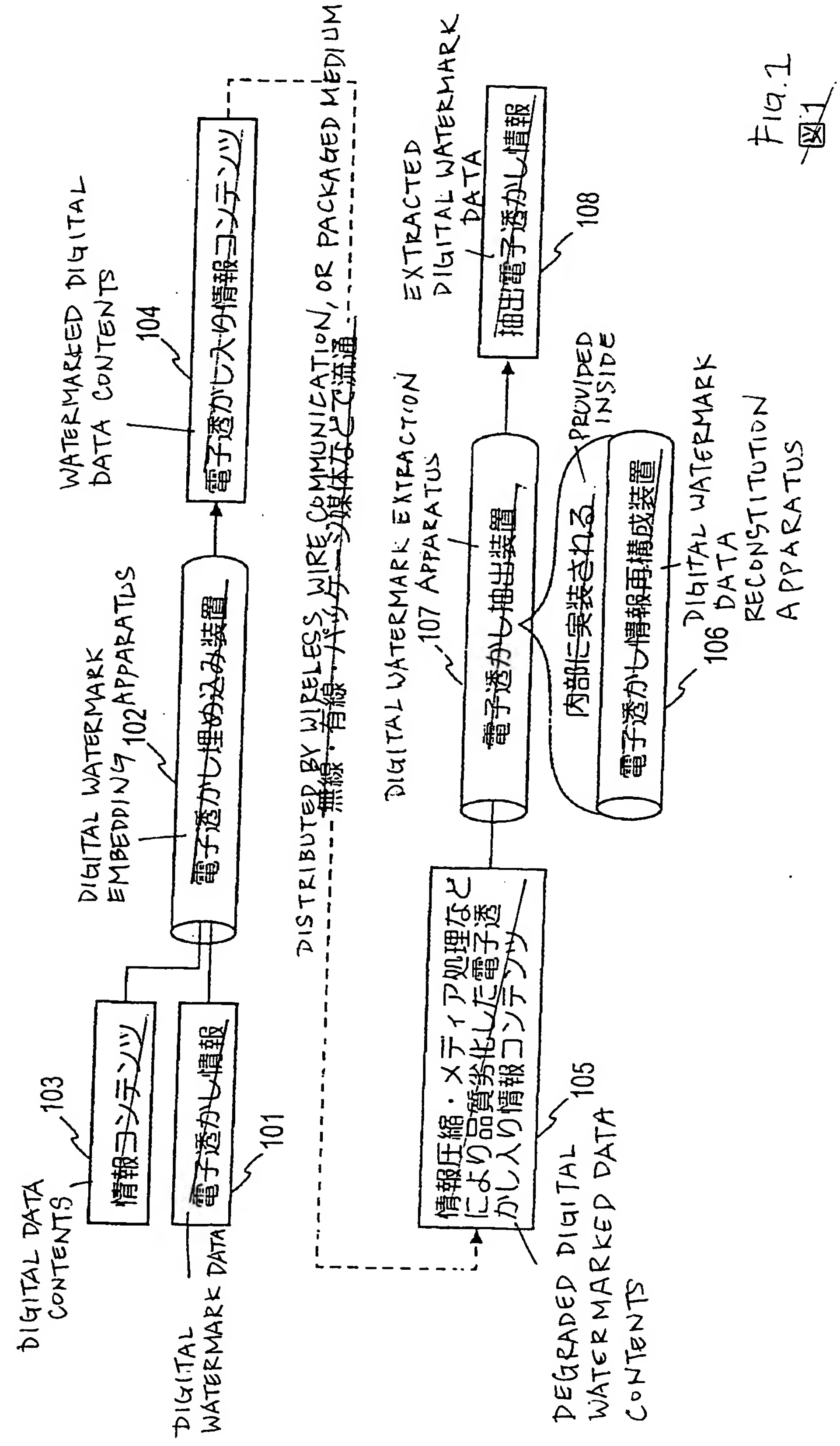
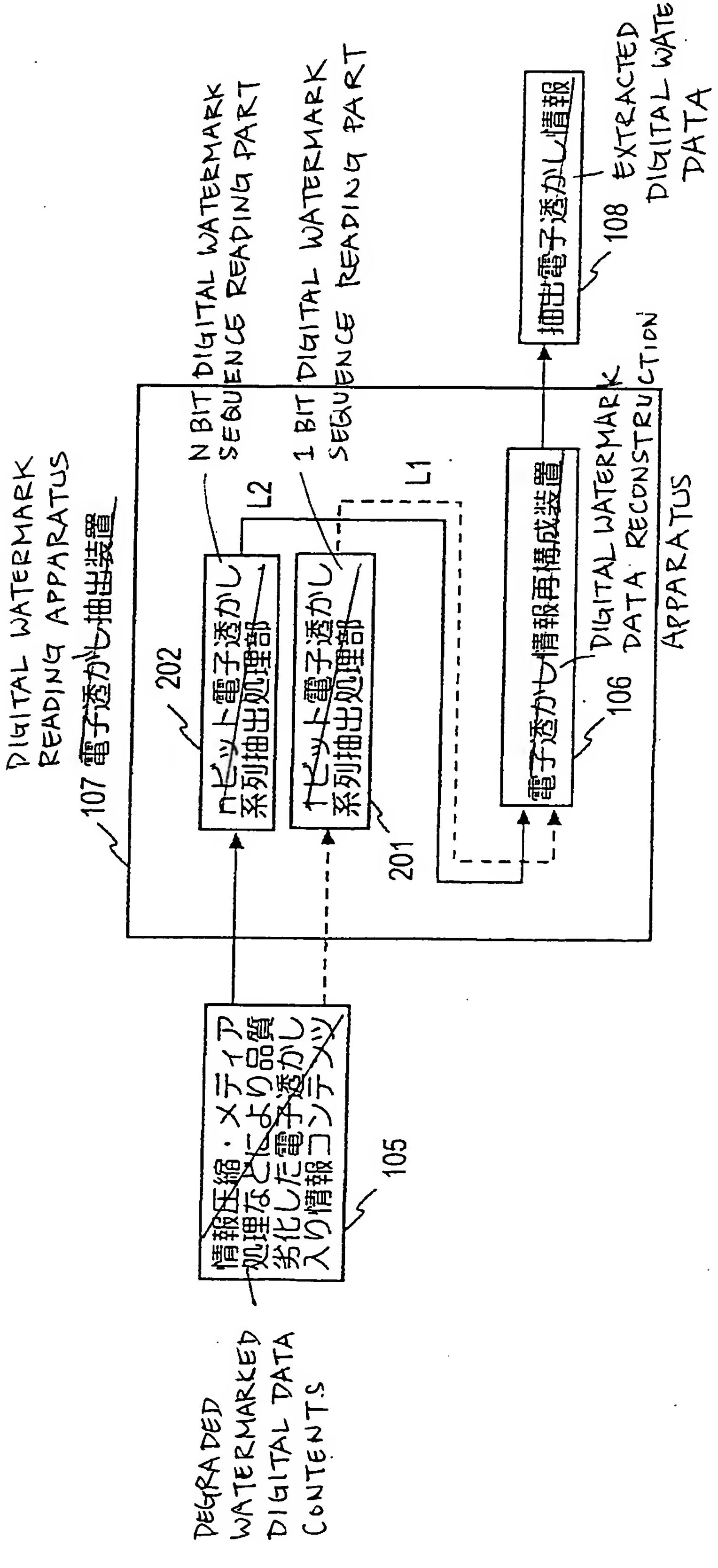

~~電子透かし情報~~
101
DIGITAL WATERMARK DATA
~~情報コンテンツ~~
103
DIGITAL DATA CONTENTS
~~電子透かし埋め込み装置~~
102
DIGITAL WATERMARK EMBEDDING APPARATUS
~~電子透かし入り情報コンテンツ~~
104
WATERMARKED DIGITAL DATA CONTENTS
~~無線・有線・パッケージ媒体などで流通~~
DISTRIBUTED BY WIRELESS, WIRE COMMUNICATION, OR PACKAGED MEDIUM
~~情報圧縮・メディア処理などにより品質劣化した電子透かし入り情報コンテンツ~~
105
DEGRADED DIGITAL WATERMARKED DATA CONTENTS
~~電子透かし抽出装置~~
107
DIGITAL WATERMARK EXTRACTION APPARATUS
~~内部に実装される~~
PROVIDED INSIDE
~~電子透かし情報再構成装置~~
106
DIGITAL WATERMARK DATA RECONSTITUTION APPARATUS
~~抽出電子透かし情報~~
108
EXTRACTED DIGITAL WATERMARK DATA

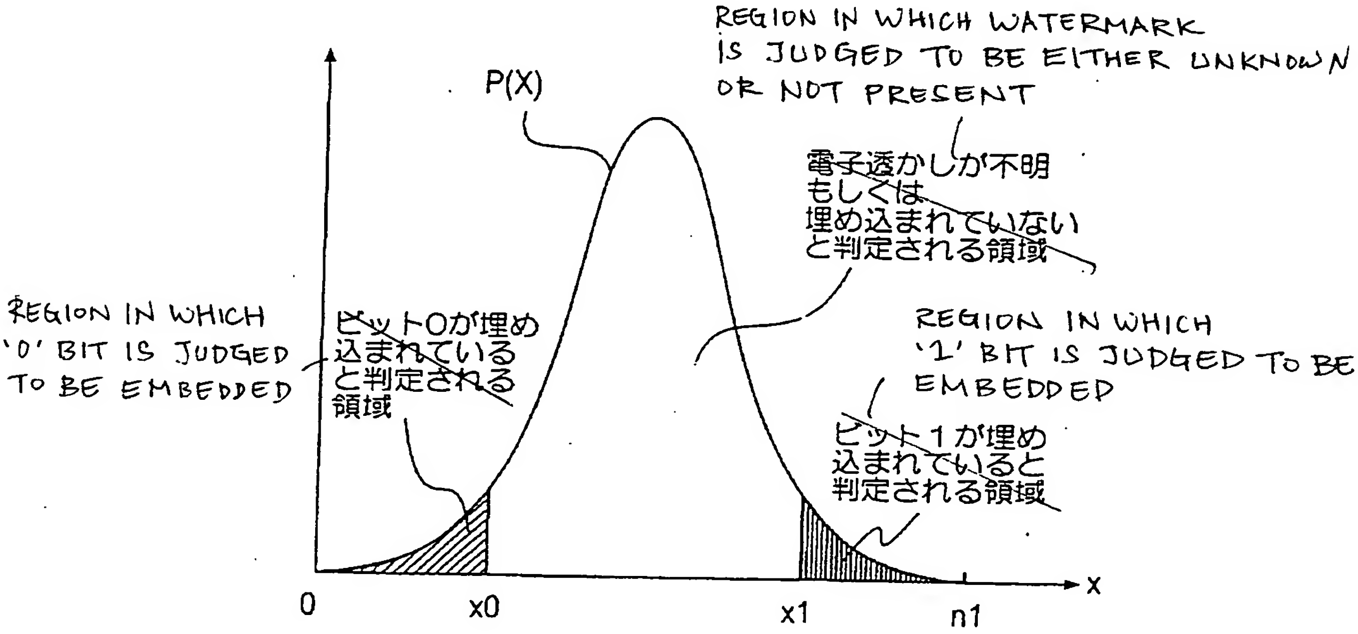
FIG. 1 ~~図1~~

【~~図2~~】FIG.2


~~情報圧縮・メディア処理などにより品質劣化した電子透かし入り情報コンテンツ~~
DEGRADED WATERMARKED DIGITAL DATA CONTENTS
105
107 ~~電子透かし抽出装置~~
DIGITAL WATERMARK READING APPARATUS
202
~~nビット電子透かし系列抽出処理部~~
N BIT DIGITAL WATERMARK SEQUENCE READING PART
201
~~1ビット電子透かし系列抽出処理部~~
1 BIT DIGITAL WATERMARK SEQUENCE READING PART
L2
L1
106
~~電子透かし情報再構成装置~~
DIGITAL WATERMARK DATA RECONSTRUCTION APPARATUS
108
~~抽出電子透かし情報~~
EXTRACTED DIGITAL WATERMARK DATA

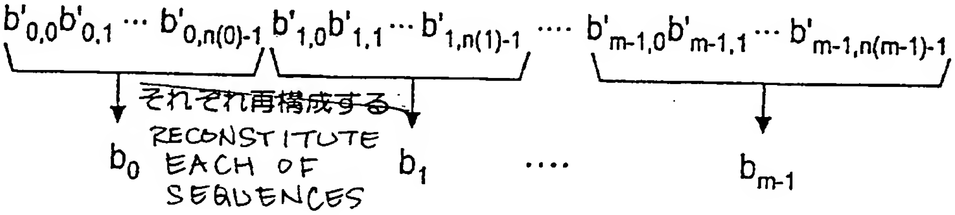
~~図2~~ FIG.2

FIG.3 ~~【図3】~~



~~図3~~ FIG.3

FIG.4 ~~【図4】~~


$b'_{0,0}b'_{0,1} \cdots b'_{0,n(0)-1}$ $b'_{1,0}b'_{1,1} \cdots b'_{1,n(1)-1}$ $\cdots$ $b'_{m-1,0}b'_{m-1,1} \cdots b'_{m-1,n(m-1)-1}$
~~それぞれ再構成する~~
RECONSTITUTE EACH OF SEQUENCES
b_0
b_1
$\cdots$
b_{m-1}

~~図4~~ FIG.4

【図5】 FIG.5

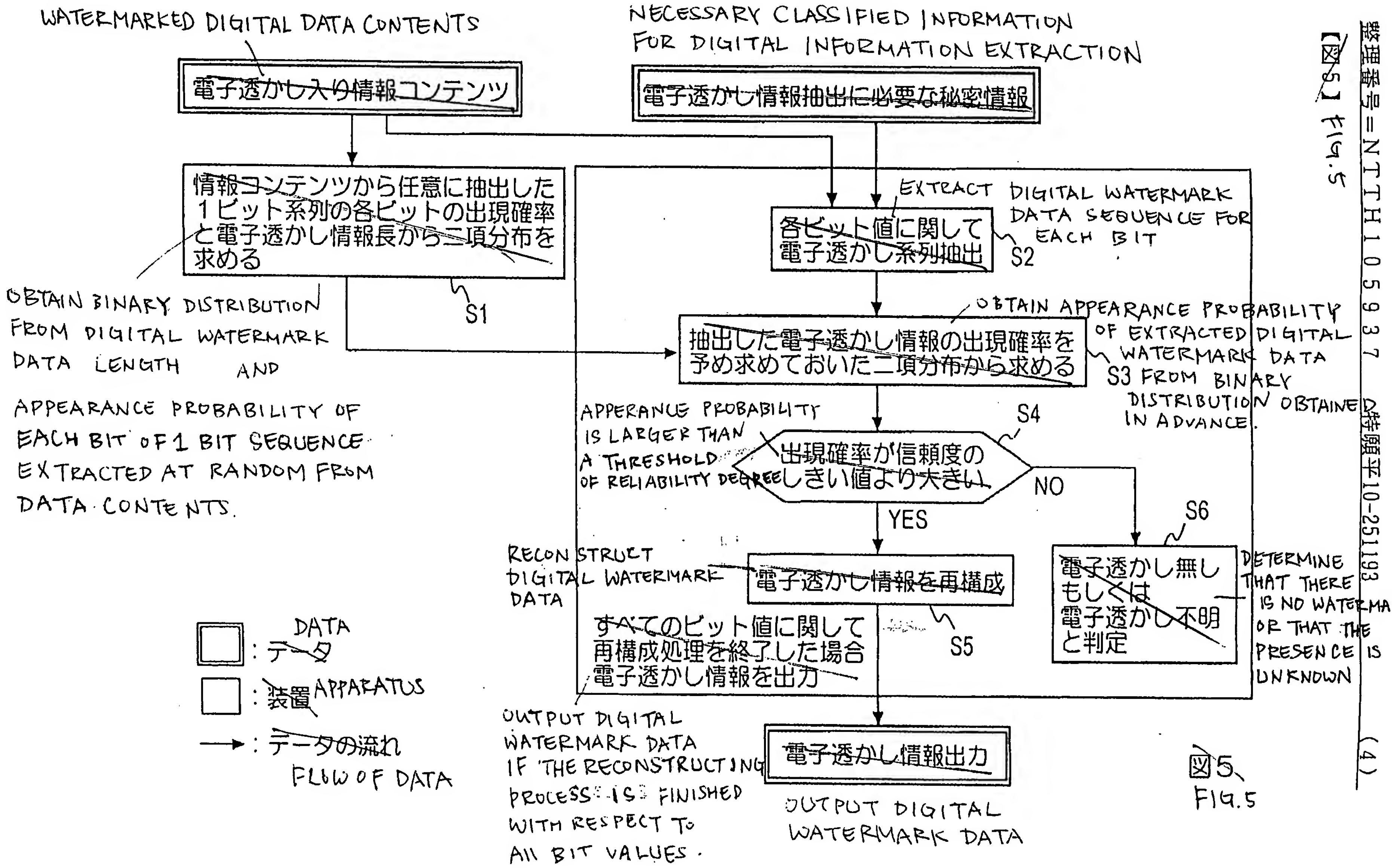
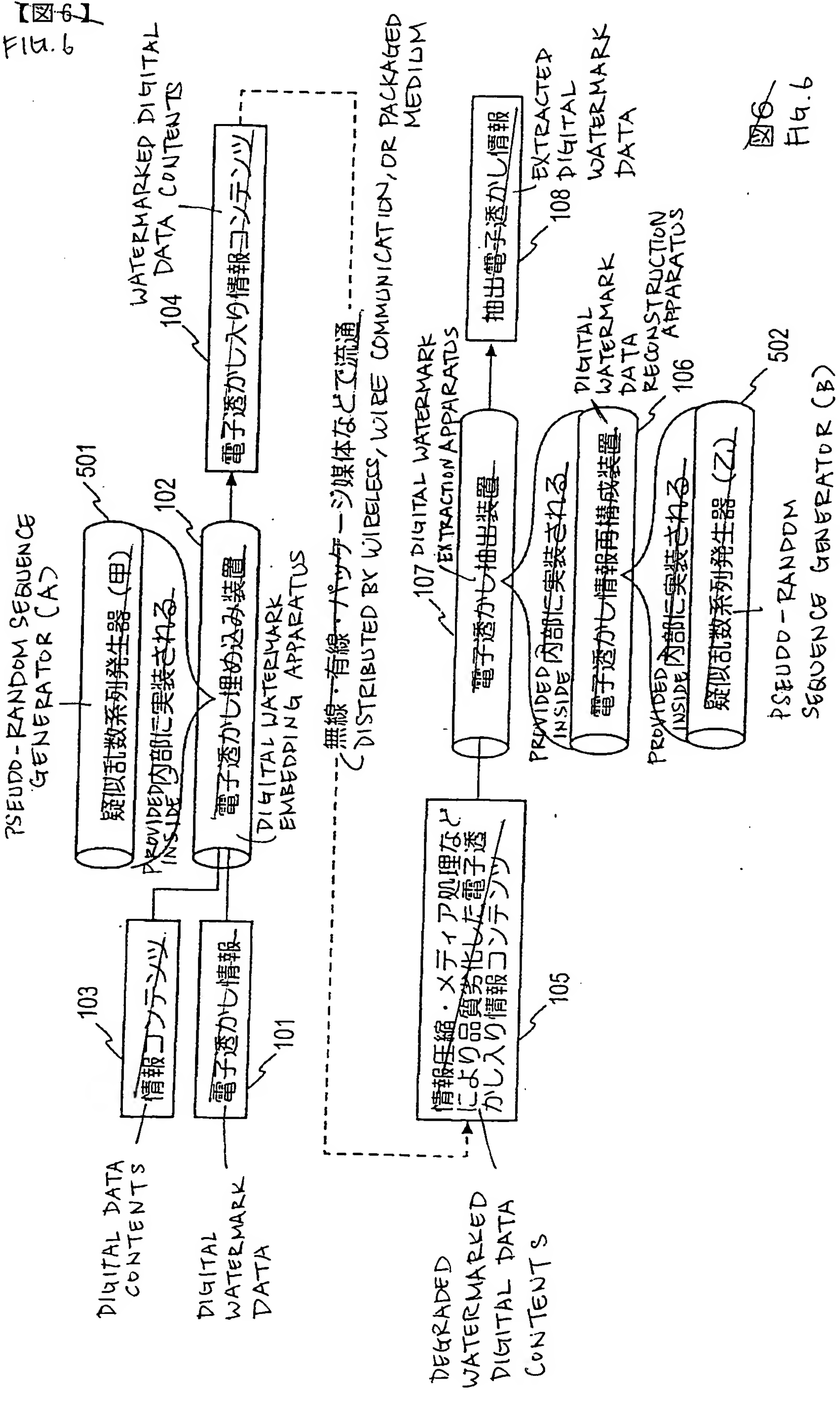

WATERMARKED DIGITAL DATA CONTENTS
~~電子透かし入り情報コンテンツ~~
NECESSARY CLASSIFIED INFORMATION FOR DIGITAL INFORMATION EXTRACTION
~~電子透かし情報抽出に必要な秘密情報~~
~~情報コンテンツから任意に抽出した1ビット系列の各ビットの出現確率と電子透かし情報長から二項分布を求める~~
S1
OBTAIN BINARY DISTRIBUTION FROM DIGITAL WATERMARK DATA LENGTH AND APPEARANCE PROBABILITY OF EACH BIT OF 1 BIT SEQUENCE EXTRACTED AT RANDOM FROM DATA CONTENTS.
EXTRACT DIGITAL WATERMARK DATA SEQUENCE FOR EACH BIT
~~各ビット値に関して電子透かし系列抽出~~
S2
OBTAIN APPEARANCE PROBABILITY OF EXTRACTED DIGITAL WATERMARK DATA FROM BINARY DISTRIBUTION OBTAINED IN ADVANCE.
~~抽出した電子透かし情報の出現確率を予め求めておいた二項分布から求める~~
S3
APPERANCE PROBABILITY IS LARGER THAN A THRESHOLD OF RELIABILITY DEGREE
~~出現確率が信頼度のしきい値より大きい~~
S4
YES
NO
RECONSTRUCT DIGITAL WATERMARK DATA
~~電子透かし情報を再構成~~
S5
~~すべてのビット値に関して再構成処理を終了した場合電子透かし情報を出力~~
OUTPUT DIGITAL WATERMARK DATA IF THE RECONSTRUCTING PROCESS IS FINISHED WITH RESPECT TO ALL BIT VALUES.
S6
~~電子透かし無しもしくは電子透かし不明と判定~~
DETERMINE THAT THERE IS NO WATERMARK OR THAT THE PRESENCE IS UNKNOWN
~~電子透かし情報出力~~
OUTPUT DIGITAL WATERMARK DATA
□ : ~~データ~~ DATA
□ : ~~装置~~ APPARATUS
→ : ~~データの流れ~~ FLOW OF DATA

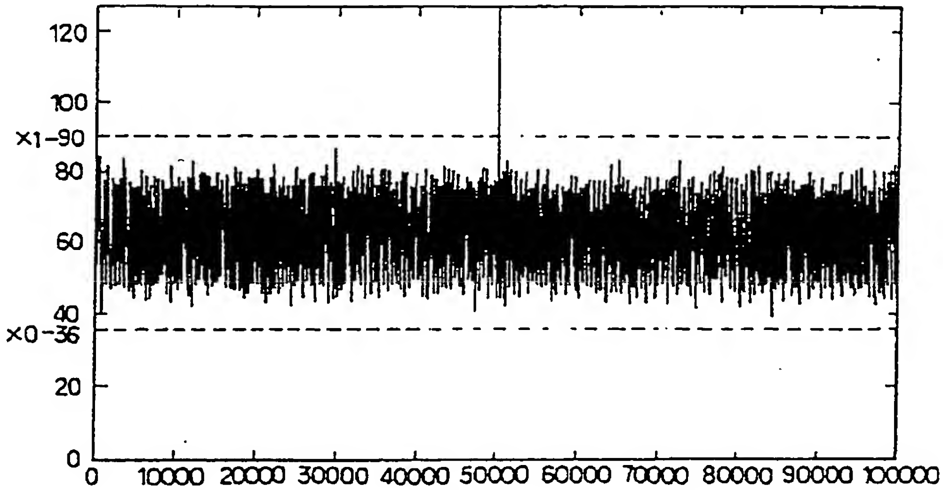
図5 FIG.5

~~【図6】~~
FIG. 6


101
電子透かし情報
DIGITAL WATERMARK DATA
103
情報コンテンツ
DIGITAL DATA CONTENTS
102
電子透かし埋め込み装置
DIGITAL WATERMARK EMBEDDING APPARATUS
501
疑似乱数系列発生器（甲）
PSEUDO-RANDOM SEQUENCE GENERATOR (A)
内部に実装される
PROVIDED INSIDE
104
電子透かし入り情報コンテンツ
WATERMARKED DIGITAL DATA CONTENTS
無線・有線・パッケージ媒体などで流通
DISTRIBUTED BY WIRELESS, WIRE COMMUNICATION, OR PACKAGED MEDIUM
105
情報圧縮・メディア処理などにより品質劣化した電子透かし入り情報コンテンツ
DEGRADED WATERMARKED DIGITAL DATA CONTENTS
107
電子透かし抽出装置
DIGITAL WATERMARK EXTRACTION APPARATUS
内部に実装される
PROVIDED INSIDE
106
電子透かし情報再構成装置
DIGITAL WATERMARK DATA RECONSTRUCTION APPARATUS
内部に実装される
PROVIDED INSIDE
502
疑似乱数系列発生器（乙）
PSEUDO-RANDOM SEQUENCE GENERATOR (B)
108
抽出電子透かし情報
EXTRACTED DIGITAL WATERMARK DATA

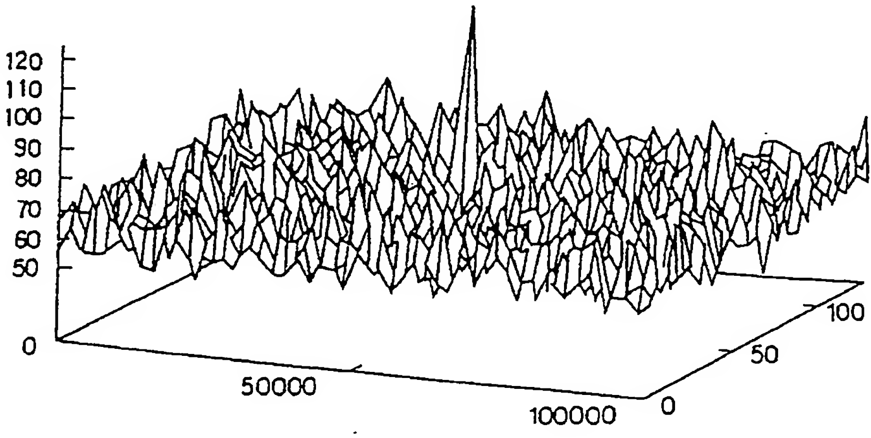
~~図6~~
FIG. 6

【~~図7~~】 FIG.7



~~図7~~ FIG.7

【~~図8~~】 FIG.8



~~図8~~
FIG. 8